

Library Consortia as Stakeholders in the Future of Bibliographic Control

**Lizanne Payne
Executive Director
Washington Research Library Consortium
901 Commerce Drive
Upper Marlboro, MD 20774
301-390-2031
payne@wrlc.org**

Library consortia are perhaps not the most obvious stakeholders in discussions of the future of bibliographic control. It is fairly easy to see the interests that individual libraries would have, and vendors, and global service providers such as OCLC. But where do consortia fit in?

First let me describe the various types of library consortia and the different metadata interests they might have. There are over 200 library consortia which have identified themselves to the International Coalition of Library Consortia (known as ICOLC). About two-thirds of these are from North America, and the rest are from every other part of the world, from Botswana to Turkey.

Library consortia provide one or more of the following services to their members:

- Joint licensing of electronic resources
- Reciprocal borrowing or shared collections, often including shared systems
- Shared library storage facilities
- Digitization projects

By far the largest number of consortia provide joint licensing of electronic resources. For many consortia that is the only, or certainly the primary, service. Examples of joint licensing consortia are the Northeast Research Libraries Consortium (NERL), the Virtual Library of Virginia (VIVA), and the Council of Australian University Librarians (CAUL).

For consortia of this type, the relationship among members is primarily financial, to contain the costs of electronic content through consortial purchasing.

Consortia which offer this service are interested in Electronic Resource Management Systems (ERMS) and metadata about licensing terms and rights management. They have a particular interest in connecting ILL rights -- that hopefully are granted in the license -- to the ILL workflow so that libraries may provide an electronic copy of articles instead of a scanned non-returnable copy. They are also interested in tracking title changes among aggregators to secure ongoing rights.

Licensing consortia are also interested in usage counts at the database and title level, to assist with renewal decisions and cost-benefit analysis.

Probably the oldest and still a very common service of consortia is support for reciprocal borrowing.

In these cases the member relationship includes sharing library collections. In these consortia, the emphasis is on metadata to support discovery, retrieval, and lending of owned items (using bibliographic, item, and patron metadata). These consortia usually share a separate common OPAC, and sometimes share a common integrated library system and other digital library systems as well. Consortia of this type with a separate OPAC include Ohiolink, and the ORBIS Cascade Alliance in the Pacific Northwest. The Washington Research Library Consortium, the Five Colleges (Ohio), the Five Colleges (MA), and the Louis consortium in Louisiana are examples of the even more tightly-integrated consortia which share an ILS as well.

Another type of consortium is one which shares library storage facilities. (And keep in mind that many consortia offer multiple programs and can belong to all of these categories, as WRLC does.)

The ReCAP facility shared by Columbia, the New York Public Library, and Princeton University, and the PASCAL facility shared by libraries in Colorado, represent cases where a separate storage consortium was formed where there was no consortial organization originally. WRLC and the Five Colleges consortia are examples of organizations which share a storage facility as well as a digital library system.

In a storage consortium, the emphasis is on metadata to support inventory control and retrieval. The inventory control system may be separate from the OPAC, or the OPAC system may perform this function.

Lastly, consortia support collaborative digitization projects, including mass digitization of library collections. The CIC consortium (Committee on Institutional Cooperation) has joined the Google Books project and will digitize up to 10 million volumes, with a specific goal to not duplicate volumes already scanned by other participants. The CIC will operate a digital repository of their scanned volumes, which has implications for linking the bibliographic information for the print and electronic versions.

Library consortia with shared catalogs and shared storage facilities are the ones which will be most directly affected by decisions about the future of bibliographic control.

Shared catalogs and systems date from a time when hardware and software were expensive and common interfaces were unavailable or unreliable (like Z39.50). Shared systems consortia will continue to play an important role because they provide a cost-effective way to support expensive IT staff and a broad range of digital library services, especially in an open-source software environment.

These shared systems are at an organizational level between the local individual library and national or global systems such as LC and OCLC.

For consortia with a separate shared catalog, data is replicated at the consortium level to facilitate searching. The economic costs are

- An extra system at the consortium level
- Loss of customized bibliographic records, which are maintained separately at the local level

For consortia where members share a common ILS, the economic costs are

- The same loss of customized bibliographic records if they share a master bib record,
- Or record redundancy if bib records are replicated for each owner. This may also be a cost to the users if they must peruse multiple bib records.

Next-generation systems with a separate presentation layer may change that configuration by allowing for separate local systems with a common user interface. This would be a variation of the “separate catalog” model, using more efficient uploaded data, not necessarily full MARC, including metadata from other sources like digital collections and institutional repositories. The Endeca implementation planned for the Florida Center for Library Automation is an example of this approach.

The primary challenge related to cataloging practices in a consortial environment is that ideally three levels of bibliographic control are needed:

- Master record for the work and/or manifestation
- Individualized cataloging as needed for consortium members, especially for subjects or other “tags”
For example, one of the reasons WRLC chose separate replicated bib records was to allow for the more detailed subject headings assigned by Catholic University. The main heading “Catholic Church” would have applied to a huge percentage of their collection.
- The third level is individual holdings for shelving and circulation data, and acquisitions

FRBR holds some promise here but is not widely implemented.

Shared storage facilities present a different set of issues related to bibliographic control.

- In high-density facilities where volumes are stored by size or in bins retrieved by robotic arms, there needs to be a place to record the actual physical location of the item, not a relative location such as call number. That is difficult to accomplish in MARC-based inventory control systems.
- Most stored items are not explicitly represented in WorldCat. That is, in most cases there is no holding library code which specifically identifies the storage facility, which means, among other things, that the storage facility does not service ILL requests directly.

Some of the more complex bibliographic control issues relate to duplicate and uniqueness detection.

A few storage facilities serve as “last copy” facilities, and others are considering it. In these cases, the facility members agree not to send duplicate materials, and share the ownership or at least have guaranteed access to the shared copy. To implement this

policy, they need algorithms to compare items before transfer to storage, to look for duplicate items. It is too slow and costly to conduct individual manual searches of items that are candidates for storage.

So duplicate detection is important. It is easy to identify some duplicate monographs by LCCN or OCLC# -- but this method is not exhaustive, and misses items which may not contain those fields.

Duplicate detection would be especially valuable for bound journal volumes, which provide the biggest return of reclaimed space for the least effort. However, MFHD data is notoriously inconsistent and resistant to automated matching. Here is a case where significant gains could be realized by having standardized data in these fields.

Uniqueness detection is also becoming important in discussions at shared storage facilities, because there is an emerging interest in using storage facilities proactively to preserve unique items. Uniqueness is not just the opposite of duplication – it is even more difficult to identify the truly unique.

A single occurrence of a given OCLC number does not necessarily represent a unique work. Even though OCLC periodically dedups WorldCat, the local library catalogs are not necessarily deduped in the same way. Few local catalogs are FRBRized, which would help with identifying unique works or manifestations.

As Karen Coyle wrote, there is a “need for unique identifiers at every hierarchical level”.

Duplicate and uniqueness detection will become increasingly important because there is an emerging interest in a system-wide program to preserve print archives – currently known as the North American Storage Trust initiative.

This would be a network of libraries and storage facilities committing to preserve and provide access to print materials, by taking advantage of materials already in long-term storage.

It would be based on collection analysis reports to compare individual library holdings to already-stored items, which would enable participating libraries to discard local holdings by relying on stored copies. This concept was described as “virtual storage” by Paul Gherman and others in a paper presented to ACRL in 2005.

There are several specific implications for bibliographic control and other metadata related to this initiative:

- It requires the ability to accurately match duplicate works or manifestations
- There is a need to correlate the stored items to the corresponding preservation commitment (as in a registry of facilities and their commitments)
- There is a strong interest on the part of some stakeholders in recording the physical condition of stored items, especially for a dark archive. However, it would be inordinately expensive to evaluate item condition retroactively, and only a little less expensive to do so prospectively

The primary economic challenges facing this initiative are:

- The native cost to libraries and consortia to upgrade the data
- The infrastructure cost to provide the collection analysis service
- The disincentive to incur these costs for the benefit of free riders – other libraries who benefit from “virtual storage” without paying for storage facilities of their own.

This may be another case where infrastructure support for long-term preservation storage at the system level should be perceived and funded as a “public good”.

Summary

Consortia provide an important organizational level midway between local libraries and the system-wide or global level, with needs slightly different from either. In many cases they provide a significant aggregation within a given geographic area: they constitute an efficient mechanism to provide service broader than the local library but more local than a national or global provider. Library consortia will watch and participate with great interest as the discussions about the future of bibliographic control proceed.